

PMCI4

EQUIVALENCE OF PAPER AND TOUCH SCREEN VERSIONS OF THE EQ-5D VISUAL ANALOG SCALE (EQ-VAS)Ramachandran S¹, Taber T², Craig BM³, Coons SJ⁴¹University of Arizona College of Pharmacy, Tucson, AZ, USA; ²Assist Technologies, Scottsdale, AZ, USA; ³University of Arizona, Tucson, AZ, USA; ⁴College of Pharmacy, Tucson, AZ, USA

OBJECTIVES: The EQ-VAS, a measure of self-reported health status, has been operationalized in ways that depart from the original format. This study examines the equivalence of the original paper-based vertical format with a touch screen-based horizontal format. **METHODS:** Non-probability sampling was employed to recruit 314 subjects intended to reflect the primary socio-demographic characteristics of the general adult population. A two part questionnaire was administered in a randomized crossover design. One part was the original paper-based 20cm vertical EQ-VAS; the other part was touch screen computer-based (designed by Assist Technologies) and included, among other items/scales, a horizontal EQ-VAS, the SF-36, and socio-demographic items. The two EQ-VAS formats were completed roughly ten minutes apart. To test for minimally important differences (MID) between EQ-VAS scores, a difference of half a standard deviation (~8 points on the 100 point scale) was used as the equivalence threshold. **RESULTS:** The mean (SD) EQ-VAS score was 81.0 (15.4) on the paper and 79.6 (15.2) on the touch-screen. The mean (CI) difference between scores on the two formats was 1.4 (0.19 to 2.58) points and the mean absolute difference was 5.3 (4.22 to 6.44) points. The intraclass correlation coefficient (ICC) was 0.75, indicating good agreement between the two scores. Almost a third (30.1%) of the respondents reported identical scores on both formats and 80.1% of the respondents had difference scores within \pm eight points. Using nonparametric bootstrap techniques, both the mean difference and the mean absolute difference between scores on the two formats were significantly less ($p < 0.001$) than the equivalence threshold. In addition, data collected via touch screen may be more reliable since 22% of subjects did not complete the EQ-VAS paper format as instructed. **CONCLUSION:** These results provide evidence for the measurement equivalence of this EQ-VAS touch screen format with the original paper format.

PMCI5

INFLUENCE OF VARIOUS HEALTH STATUS MEASURES ON TOTAL EXPENDITURES IN THE MEPS DATASETDe Smet BD¹, Erickson S²¹The University of Michigan Health System, Ann Arbor, MI, USA;²University of Michigan, Ann Arbor, MI, USA

OBJECTIVES: The Model of Health Services Use is an analytic framework incorporating predisposing, enabling, and need predictor variables to explain patients' use of health care resources. We applied this model to the consolidated year 2000 Medical Expenditure Panel Survey (MEPS), a representative survey of the US civilian, non-institutionalized population, to compare the influence of various need variables, perceived health status, on the outcome of total health-related expenditures. **METHODS:** Multivariate linear regression models were developed, maintaining a core set of predictor (predisposing and enabling) and dependent (total health care expenditures) variables, varying only the need predictor variables. Predisposing variables include age, gender, race, education, and marital status; enabling variables include insurance type, employment status, family size, and household income; need variables included summary scores of two general health status measures (SF-12 PCS and MCS and the EQ-5D—Index and VAS), and single-item core MEPS questions of perceived mental and physical health status. Analysis took

into account the complex design of the dataset. The R^2 of each model is presented for descriptive comparison. **RESULTS:** Data from this MEPS dataset was obtained from 16076 respondents, representing over 209 million US residents. The base model, containing only predisposing and enabling variables, had an R^2 of 0.064. The models using the single-item core questions of perceived mental and physical health status separately yielded R^2 values of 0.073 and 0.104, respectively, with an R^2 of 0.105 when both were included. R^2 values for the models containing the SF-12 MCS, SF-12 PCS, EQ-5D index, and EQ-5D VAS individually were 0.068, 0.084, 0.065, and 0.071, respectively. Including the SF-12 MCS and PCS together in one model, the R^2 was 0.086. **CONCLUSION:** The results imply that the single-item core health-status questions used by the MEPS perform marginally better than the SF-12 or EQ-5D to explain total expenditures.

PMCI6

LINKING DISEASE-SPECIFIC QUALITY OF LIFE (QOL) SCALESMcKenna SP¹, Meads DM¹, Doward LC¹, Tennant A²¹Galen Research, Manchester, UK; ²University of Leeds, Leeds, UK

OBJECTIVES: To identify the most effective method of linking disease-specific scales through the application of Rasch analysis. **METHODS:** Scales assessing rheumatoid arthritis (RAQoL) and adult growth hormone deficiency (QoL-AGHDA) were selected for linkage. Interviews were conducted with 38 patients to identify additional items that were relevant to both diseases. A postal survey was then conducted with 103 RA and 98 GHD patients. Two main linking approaches were assessed; linking the two scales by the nine additional (common) items identified and use of an independent anchor or test (the PGWB). Here, all items in the PGWB are combined with all items in each of the scales. **RESULTS:** Adding the nine common items identified to the RAQoL led to a scale with excellent fit to the Rasch model; Item Fit (mean = -0.19, SD = 1.22), Person Fit (mean = -0.21, SD = 0.89) and Person Separation Index (0.94). Adding the nine items to the QoL-AGHDA also led to excellent fit to the model; Item Fit (mean = -0.14, SD = 1.46), Person Fit (mean = 0.015, SD = 0.84) and Person Separation Index (0.96). Comparison of scores on the nine common items suggested that the GHD group had worse quality of life than RA patients. Use of the PGWB as a linking test led to considerable item misfit in both scales. **CONCLUSION:** Use of the PGWB as an anchor test was unsuccessful (probably as it assesses well-being (impairment) rather than QoL). For the purposes of constructing an item bank common item equating appears to be feasible. Such co-calibration provides an opportunity for valid and accurate comparisons of the impact of different diseases on patient groups. It must be noted that co-calibration requires that the scales to be linked adopt the same measurement model.

PMCI7

THE UK EQ-5D_{INDEX}: AN EVALUATION OF FACE VALIDITY IN HOSPITAL TREATED SUBJECTS

Currie CJ, McEwan P

Cardiff University, Cardiff, UK

OBJECTIVE: The EQ5D_{index} is widely used to evaluate health preferences and provide utility estimates. Objective of this study was to evaluate the face-validity of the EQ5D_{index}. **METHODS:** Data used here were the first 40,000 responses in the Health Outcomes Data Repository (HODaR). In addition to survey data HODaR details clinical phenotype. Patients were surveyed with the EQ5D_{index} (excluding the VAS), and the SF36. A simultaneous measure quantified health status on an arbitrary scale of 0 to 100 pre-admission and post-discharge. The EQ5D_{index} has 243

potential values. **RESULTS:** Frequency distribution of the EQ5D_{index} was tri-modal and difficult to describe in summary statistics. In all, 27 possible values (11%) were responsible for 92% of all observations, 14 possible values had no observations, and 24.7% of returns had an EQ5D_{index} of 1.0. There are a number of categories that are rarely used e.g., severe mobility problems and severe self care problems. There was a close correlation between weighted scale and simple addition of responses ($R^2 = 0.87$). There were 6.8% of responses with an EQ-5D_{index} ≤ 0.0 . There was a low correlation between the EQ5D_{index} with the general health question of the SF36 and the arbitrary, continuous valuation of health status above. The ranking of mean estimates was intuitively correct. **CONCLUSIONS:** The number of theoretical values that are represented was sparse. The EQ-5D_{index} distribution results in no easily describable parametric distribution, and the correlation with other general health measures was low. Given that these subjects are hospital treated, too many may have a health status of 1.0, and too many are also in a health status notionally equal to or worse than death. Decisions based on the EQ5D_{index} now have enormous health and commercial implications. The EQ5D classifies the right health factors but the sensitivity and scoring methods need urgent revaluation: good but needs improving.

PMC18

HEALTH UTILITIES INDEX (HUI) ON-LINE QUESTIONNAIRE SYSTEM: CRITERION VALIDITY OF MULTI- AND SINGLE-ATTRIBUTE UTILITY SCORES

Hunter D, Furlong W, Horsman JR

McMaster University, Hamilton, ON, Canada

OBJECTIVES: To assess the criterion validity of HUI Mark 2 (HUI2) and Mark 3 (HUI3) utility scores from a new, centralized on-line questionnaire administration system. **METHODS:** The system presents HUI questionnaires to patients and provides results to clinicians by email. Questionnaire results include responses and 32 derived variables (14 attribute levels; two overall health state vectors; 14 single-attribute utility scores; and two multi-attribute utility scores of health-related quality of life (HRQL)). SPSS code, validated to Health Utilities Inc. decision tables for determining attribute levels and published utility functions, is the criterion method for determining HUI derived variables. Testing used a data set that included questionnaire response combinations for all HUI2 and HUI3 attribute levels. Criterion validity was evaluated using percent exact agreement, and single-measure intra-class correlation coefficients (ICC), between scores from the new system and scores from the criterion method. **RESULTS:** The test data set generated 240 utility scores. There was exact agreement for 99.2% ($n = 238$) of the scores. Disagreement was limited to HUI2 sensation (ICC = 0.805, $p < 0.01$), and HUI2 overall HRQL (ICC = 0.966, $p < 0.01$), scores in one test case. Results were received by email from the on-line system within approximately one minute of completing each questionnaire. There were no missing or incomplete questionnaire data from the on-line system. **CONCLUSION:** The results indicate that most of the utility scores from the new on-line questionnaire system have criterion validity and there is a problem with the coding algorithm for at least one set of questionnaire response combinations associated with HUI2 sensation. The coding problem should be corrected, and more rigorous testing should be completed, before public release of the system. The system should be considered an alternative to traditional methods for future HUI data collection, especially for applications requiring immediate results such as clinical settings.

IMPROVING THE SCALING PROPERTIES OF THE PSYCHOLOGICAL GENERAL WELL-BEING SCALE (PGWB)

McKenna SP¹, Meads DM¹, Doward LC¹, Tennant A²

¹Galen Research, Manchester, UK; ²University of Leeds, Leeds, UK

OBJECTIVES: To apply item response theory (IRT) to PGWB data to determine whether the instrument provides unidimensional assessment of well-being; identify a revised version of the measure. **METHODS:** The PGWB is a widely used patient-completed generic measure of well-being that, to date, has not been subjected to item response theory (IRT) assessment. The UK version of the PGWB was used. It consists of 22 items, each with five response options. The measure was completed by two patient groups, 103 patients with rheumatoid arthritis and 96 with adult growth hormone deficiency. Data were subjected to Rasch Analysis using RUMM 2010. **RESULTS:** Analysis revealed problems with the five option scoring system for four of the items. Three methods of analysis were followed to obtain the best fit of data: 1) rescoring of disordered items and deletion of any further misfitting items; 2) collapsing response options into three categories and rescoring further misfitting items; and 3) deletion of disordered items and any further misfitting items. The first approach gave the best fit of the data to the Rasch model in terms of overall and individual item fit and person-item separation. Three other items were then removed due to poor item fit. Subsequently, fit to the Rasch model was good, in terms of overall Item-Trait Interaction ($\chi^2 = 128.87$, $df = 95$, $p = 0.001$), Item Fit (mean = 0.156, SD = 1.592), Person Fit (mean = -0.294, SD = 1.297) and person Separation Index (0.955). **CONCLUSION:** Application of Rasch analysis to PGWB data identified a new version of the instrument consisting of 19 items with good scaling properties. Use of the new version would improve the accuracy with which well-being is assessed in clinical studies. It is recommended that the new version is tested with other disease groups to determine whether the scaling properties are maintained.

PMC20

RELATIONSHIP BETWEEN PATIENT SATISFACTION AND PERCEIVED HEALTH STATUS

Xiao H

Florida A&M University, Tallahassee, FL, USA

OBJECTIVE: To examine the relationship between patient satisfaction with access to care and their perceived health status. **METHODS:** Information on patient satisfaction with access to care and perceived health status along with their demographics was extracted for people 35–64 years of age, from the Household Component of 1999 Medical Expenditure Panel Survey. Descriptive statistics were used to illustrate the characteristics of the study population. Multiple regression analysis was applied to examine the relationship between patient satisfaction and their self-rated health status controlling for age, gender, race, marital status and education level. All analyses used STATA 8.0 which is designed to analyze weighted data. **RESULTS:** A total of 8746 patients met the study inclusion criteria and were included in the study. Of these patients, 53% were females, 69.3% were married and 82% were Caucasian. The majority (44.9%) had a high school diploma and 14.5%, 6.9% and 1.7% held BS, MS and Ph.D. degrees, respectively. Patients who rated their health better scored higher in their satisfaction with access to care. In addition, higher satisfaction was found in patients with the following characteristics: being older, female, Eskimo, married and with higher education. Asian and Hispanic patients scored lower in satisfaction than Caucasian patients. **CON-**